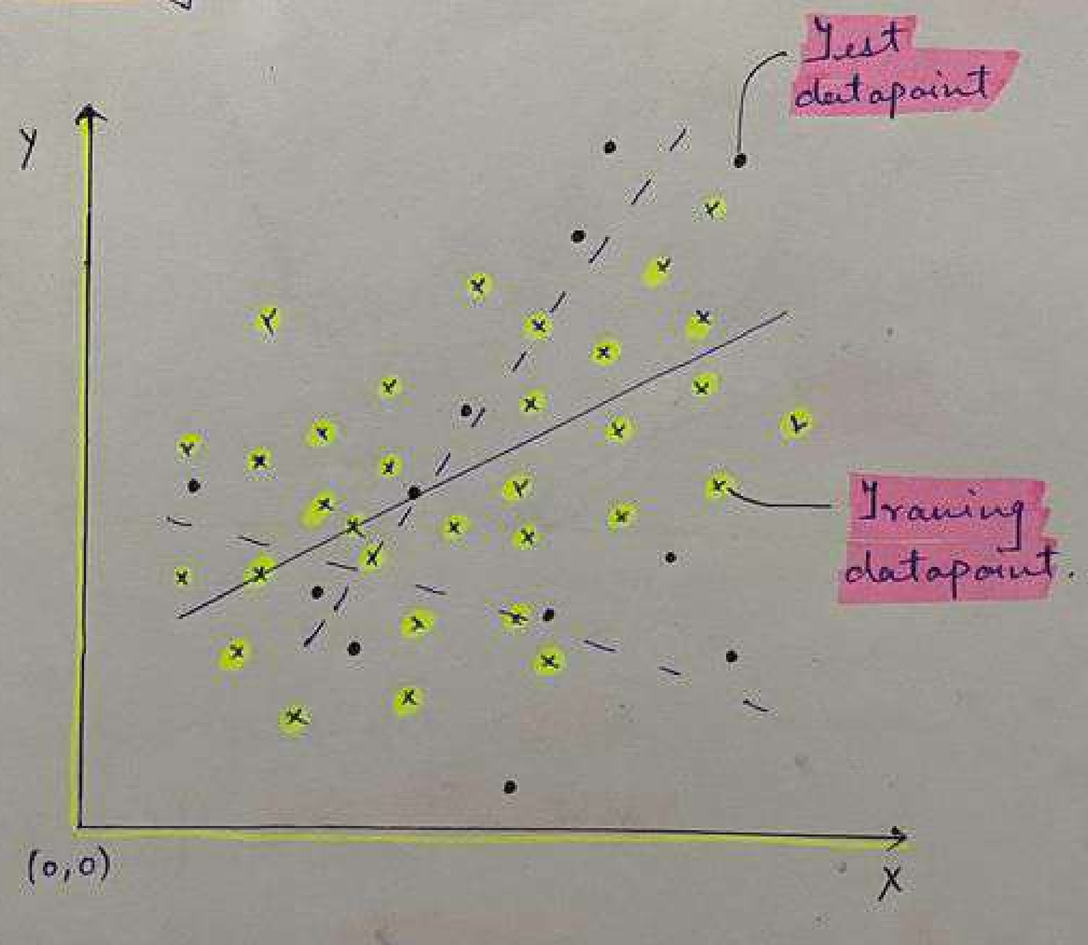
Ridge Regression

- when we have overfitting condition with our Model we uses Ridge Regression.

It's also known as L2 Regularization.

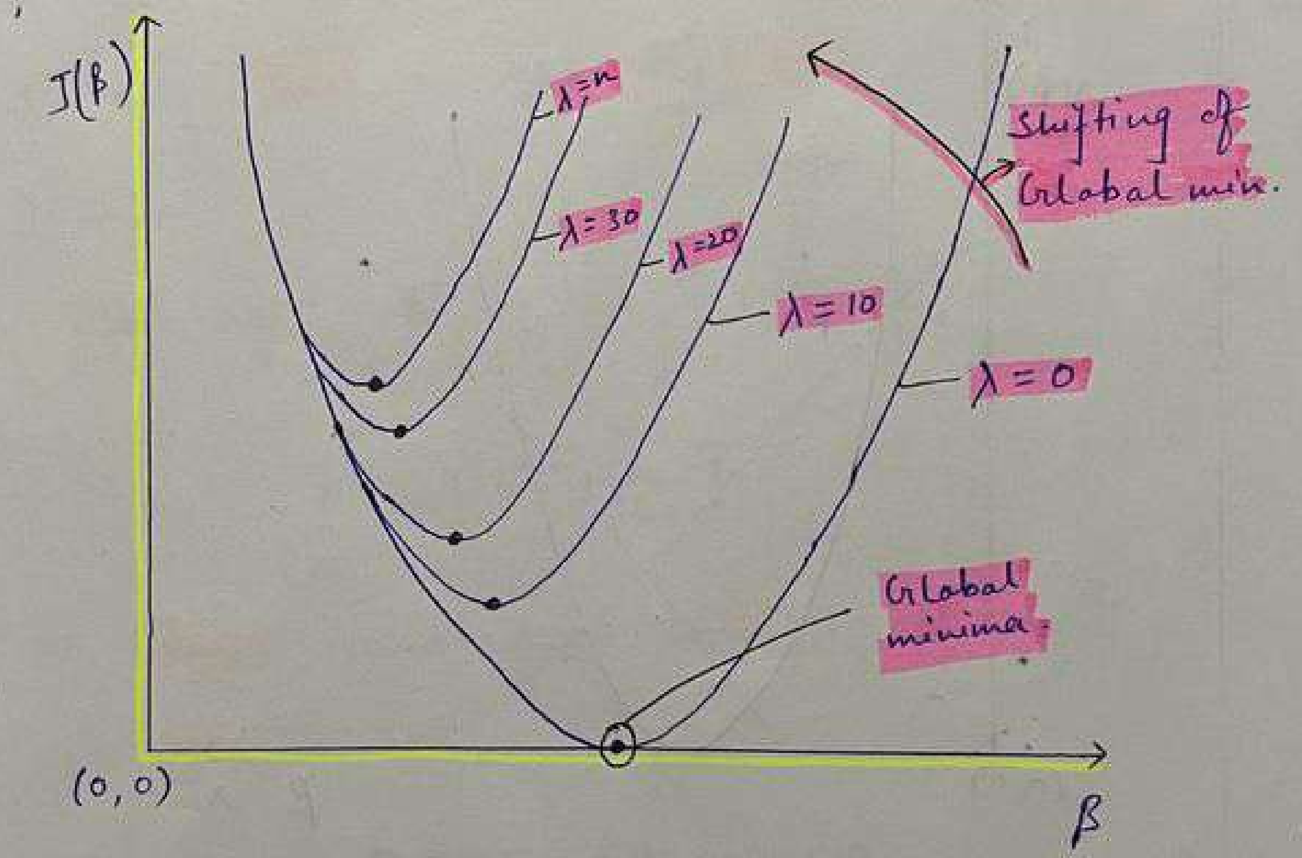
It's a model tunning method used to analyse any data that suffers from multi-collinearity.



In above case best fit line of Traning data will not fit for Test data & dotted lines will probably predict well for Test datapoints.

$$\text{Ridge reg} : \underbrace{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}_{\text{Cost function}} + \lambda \sum_{i=1}^{n}(\text{slope})^2$$

λ : Hyperparameter.



when we consider $\lambda = 0$, we are left with just cost function.

Similarly after changing value of λ the Global minimum is shifted upward. toward left mean value of slope and β is decreasing.

From above we have :

$$\lambda \propto \frac{1}{\text{slope}}.$$

let's assume for instance cost function to be zero in case if training data totally lies on the best line.

Then,

$$\text{Ridge reg} : 0 + \lambda \sum_{i=1}^{n} (\text{slope})^2$$

$$= \lambda \times \begin{bmatrix} +\text{ive} \\ \text{value} \end{bmatrix}$$

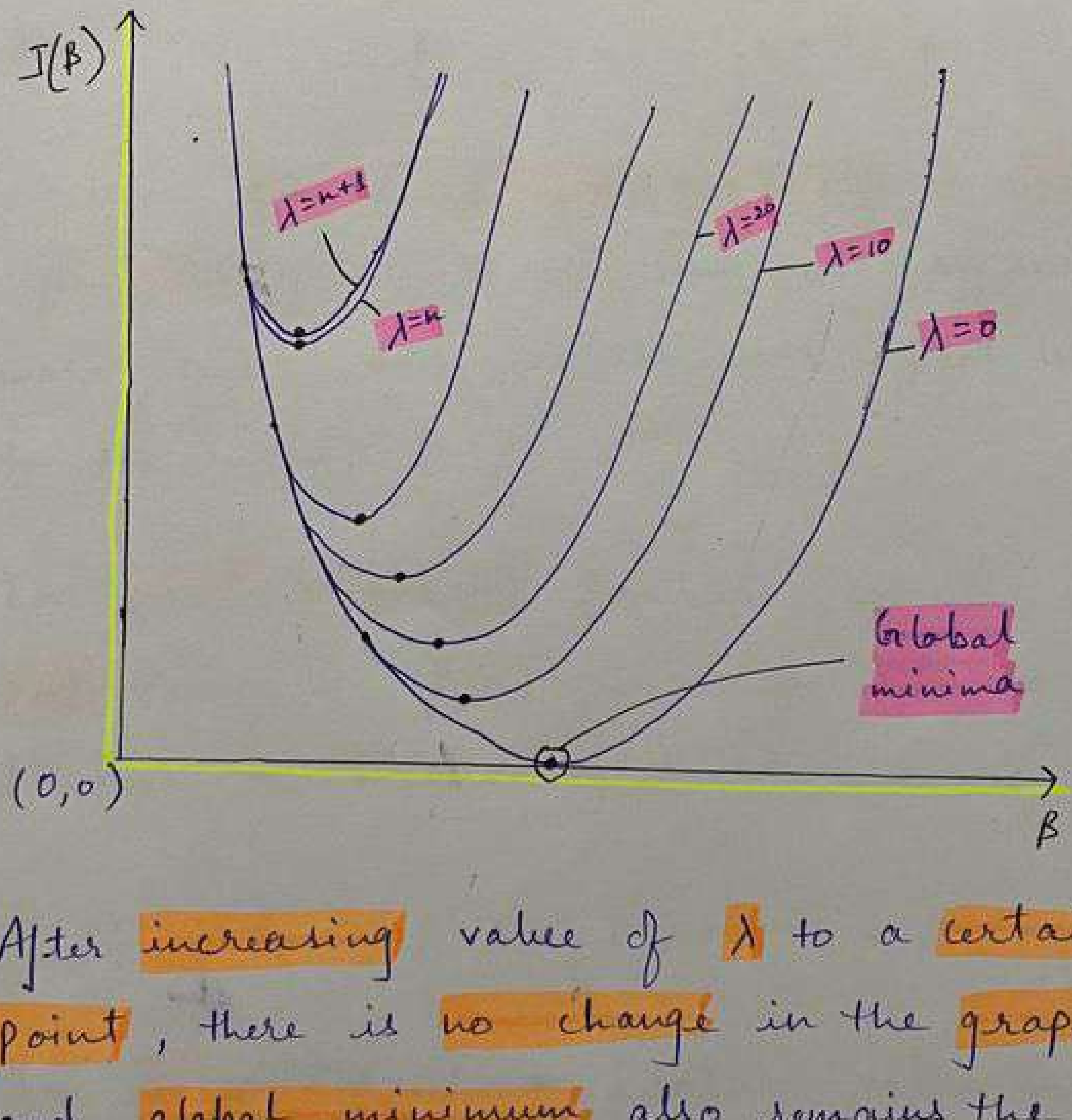
$\Rightarrow$ value will keep changing but not overlap.

Lasso Regression.

— This method uses shrinkage technique. It shrunk the value towards a central point as of mean. It's also known as L1 regularization and is used to reduce features which are not required.

$$\text{Lasso reg} : \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 + \lambda \sum_{i=1}^{n} |\text{slope}|$$

Features that are not important are eliminated using Lasso regression.



After increasing value of λ to a certain point, there is no change in the graph and global minimum also remains the same.

Consider,

$$\beta_0(x) = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \beta_3(x_3)$$

In above case we have three independent variable. Consider value of β_1, β_2 & β_3 as .5, .01, .2

Then, $\beta_0(x) = \beta_0 + .5\,x_1 + .01\,x_2 + .2\,x_3$

we have considered random value of independent features as some will affect more and some feature will affect less to the model.

For least dependent feature when we put value of $\beta_0(x)$ in slope and is further multiplied by λ it becomes.

$$\therefore \lambda\beta_0 + \lambda .5\,x_1 + \lambda \cdot .01\,x_2 + \lambda \cdot .2\,x_3$$

For least dependent feature value will decrease further after multiplication with λ.

Thus after increasing λ value after a limit slape is stucked and least dependent variable tends / approach to zero.

Elastic Net

- It's combination of L1 and L2 regularization.

$$\boxed{: \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\lambda_1\sum_{i=1}^{n}(\text{slape})^2+\lambda_2\sum_{i=1}^{n}|\text{slape}|}$$

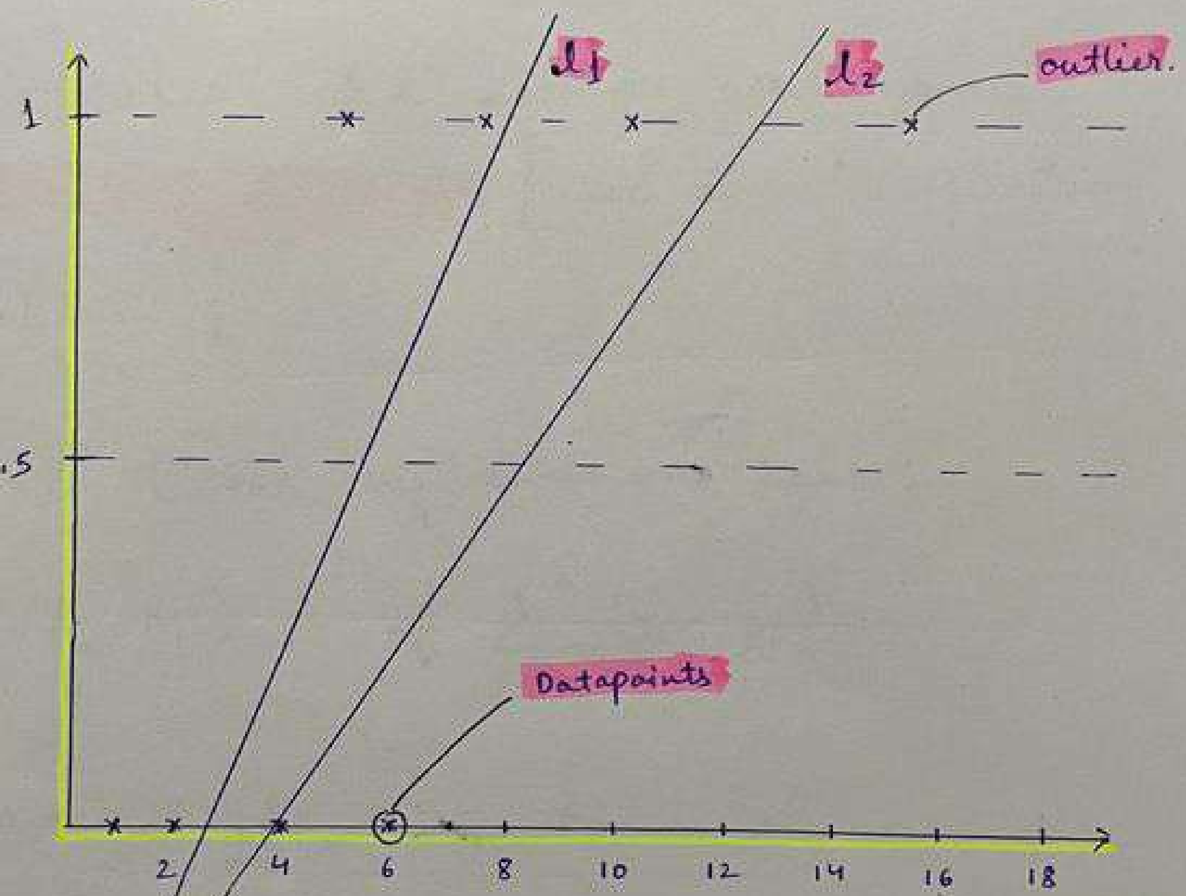
Logistic regression.

Consider below prablem statements:

Study hr	Pass/Fail.
1	Fail
2	Fail
10	Pass
6	Fail
5	Pass
7	Pass
4	Fail.

From above we can conclude Pass / Fail situation as 1/0.

Plotting same with prediction using linear regression.



L_1 shows narmal line when we have no out lier in our dataset. But L_2 represent line when we have outlier. for same one who study 15 hr. Then line will be shifted from $L_1 \rightarrow L_2$ and one even studying 10 hr will not pass & value is negative if study less than 4 hr in our new line L_2.

So we need to bound it within a range in other word we need to squash it within bounded range & then we can find best fit line.

In cases like above we use logical regression [used in classification problems]

From linear regression we have:

$$J(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

also, $y(x) = \beta_0 + \beta_1 x$

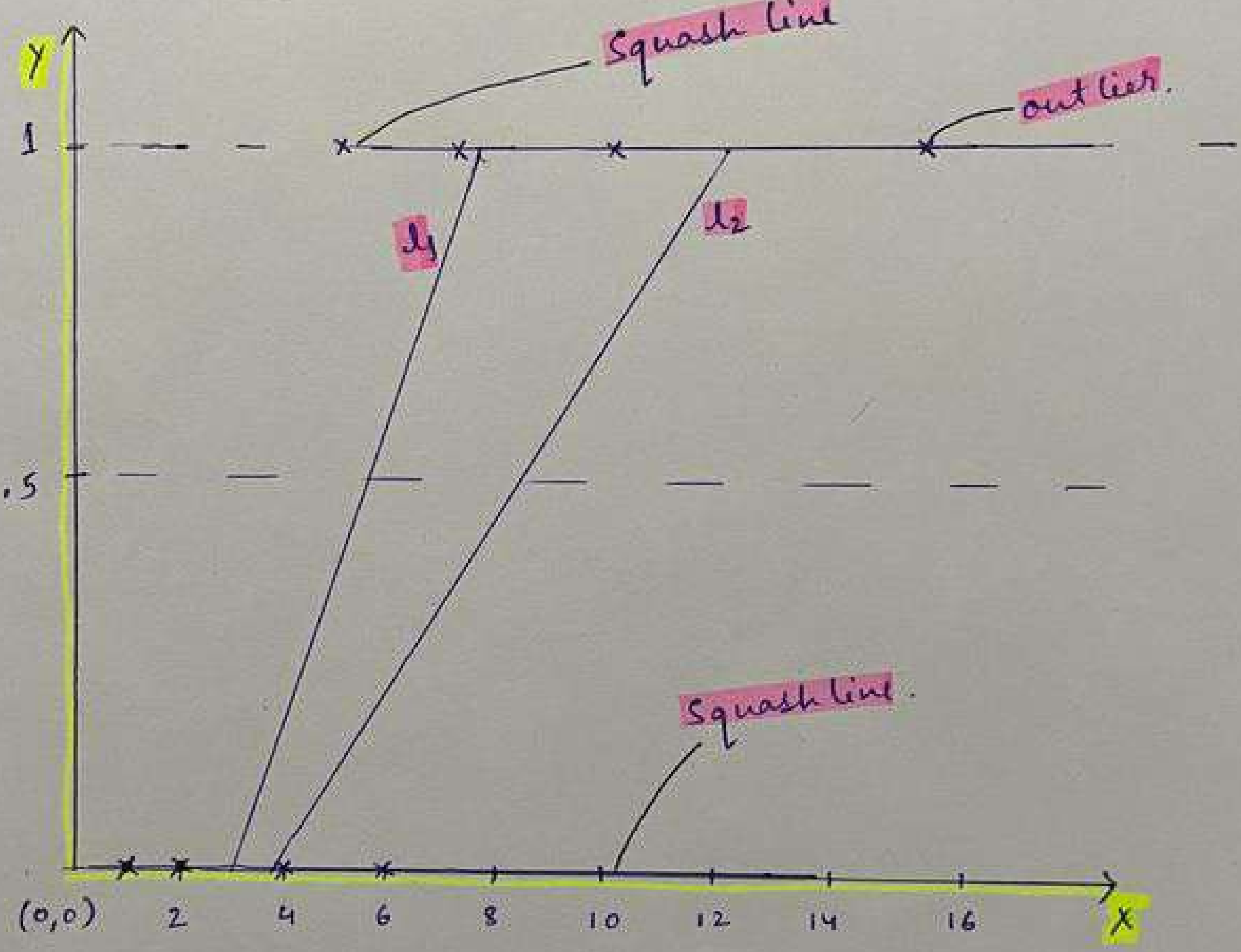
we squash above equation using sigmoid function.

$$\boxed{\text{Sigmoid function} = \frac{1}{1+e^{-z}} \quad \begin{bmatrix} z = y(x) \\ = \beta_0 + \beta_1 x \end{bmatrix}}$$

Value is now bounded b/w:

0 — 1

From above it's clear that sigmoid function is related with the cost function so follow convex function. Thus slope can be found easily.



logical regression cost function:

Sigmoid coef $= \sigma$, logical expression $= \beta_0 + \beta_1 x$

Applying sigmoid function: $[z = \beta_0 + \beta_1 x]$

$= \sigma(\beta_0 + \beta_1 x) = \sigma(z)$

$= \dfrac{1}{1 + e^{-z}} = y(x)$ for logical reg.